The metatheory of first-order logic: a contribution to a defence of *Principia Mathematica*

Stephen Boyce
University of Sydney

29 November 2010

Abstract

This paper presents an account of the first-order logic of *Principia Mathematica* and evidence that the system is superior to currently accepted classical rivals. A widely accepted view that Whitehead and Russell's presentation of logic in *Principia* is essentially defective in departing from contemporary standards is considered. It is shown that the judgement is based on a number of arguments that do not withstand critical scrutiny. For example, the assumption that the metatheory of contemporary first-order logic may be made precise in the expected way (using a first-order set theory equivalent to NBG) is shown to be false; on pain of contradiction, there cannot exist any such domain of interpretation of NBG. An alternative view of first-order logic, derived from *Principia*, is then presented. It is shown that *Principia* avoids the problem just discussed, as the first-order fragment may be made precise under an interpretation of the full system.

1 Is there a logic of *Principia*?

My primary aim in the following is to sketch an account of the first-order logic of *Principia Mathematica*. I begin by considering the claim that the Whitehead-Russell presentation is flawed in so far as modern standards for the characterisation of a system of logic are not adhered to; hence to recover a system of logic from *Principia* it is essential to consider a revised system that adheres to contemporary standards must be developed. I challenge this claim below (§2.1) by sketching a proof that the metatheory of modern first-order logic cannot be made precise in the generally expected way; more precisely, I show that (Proposition 2.1) there is no interpretation of NBG set theory that expresses the metatheory of Mendelson's pure predicate calculus *PP* ([23]) .

In §2.1 I focus on a standard presentation of a contemporary first-order logic, rather than a modern reconstruction of the first-order fragment of *Principia*, in order to support the claim that the Whitehead-Russell account of first-order logic ought to be considered on its merit. As the Whitehead-Russell account has been dismissed by many highly regarded authors (§2) it is necessary to present evidence in support of this claim. Having presented such evidence, I then (§3) examine the Whitehead-Russell version of first-order logic. I focus on the question of how to make *Principia*'s notions of a proposition / propositional functions precise - as a widely accepted objection to *Principia* holds that some confusion (perhaps involving sloppiness about the use/mention distinction) is intrinsic to the Whitehead-Russell account (cf [26], [3]: 56, [18]: 259). The slogan version of my account is that the logic of *Principia* is a logic of propositional functions, the latter being neither open sentences nor the 'meanings' attached to these under some interpretation(s) but rather a combination of the two. In §5 I sketch an extension of this account to the full system of *Principia*.

For brevity throughout the following various well-known distinctions concerning formal systems are blurred where no confusion should result. For example, it is convenient in many contexts to speak of the symbols of a formal system when it is really instances of the symbols that are mentioned; similarly, it is often convenient to refer to the meaning of a sentence and so on, when it is really the meaning of the sentence in a specific context that is referred to. As such distinctions are well-known in the literature and nothing essential rides on them in what follows I will make no further mention of them herein.

2 The formalist drive-by

In this section I wish to consider the following claim (hereon 'the formalist drive-by'): to recover a system of logic from *Principia* it is essential to abandon the Whitehead-Russell presentation and construct an alternative system that adheres to contemporary standards. Urquhart, in describing the type theory of *Principia*, explicitly asserts this claim:

> ... it is better to abandon the original Whitehead-Russell presentation of the ramified theory of types and to follow a modern

> presentation. The original presentation in *Principia Mathematica* is both imprecise and notationally clumsy. Above all, the original formulation is unsatisfactory because there is no precise presentation of the syntax of the system. [29]: 295

The formalist drive-by is asserted implicitly by some authors. Gödel, for example, to establish his famous incompleteness theorems concerning systems such as *Principia*, explicitly examines the formalist system P rather than the system of logic that Whitehead and Russell actually present in *Principia*. Gödel is keen in [11] to establish that the exhibited proof catches the system of *Principia* itself; not surprisingly therefore the differences between the two systems are asserted to be of no essential import:

> P is essentially the system obtained when the logic of *PM* is superposed upon the Peano axioms ... The introduction of the Peano axioms, as well as all other modifications introduced in the system *PM*, merely serves to simplify the proof and is dispensable in principle. [11]: 599

That Gödel actually judged the formalist reconstruction of, for example, *Principia*'s inference rules to be an essential correction to the Whitehead-Russell presentation is however made clear elsewhere. In characterising *Principia*'s place within the development of logic more generally Gödel states:

> It is to be regretted that this first comprehensive and thoroughgoing presentation of a mathematical logic and the derivation of mathematics from it [is] so greatly lacking in formal precision in the foundations (contained in *1 - *21 of *Principia*) that it presents in this respect a considerable step backwards as compared with Frege. What is missing, above all, is a precise statement of the syntax of the formalism. ([12]: 120, modified through interpolation '[]')

Gödel's considered position is thus that to extract a system of logic from *Principia* the Whitehead-Russell presentation must be modified so that a system with a precisely characterised, formal syntax (and so on) is described. When implicit and explicit assertions of the formalist drive-by are grouped together, it is clear that this claim is widely endorsed by many highly regarded commentators: [4], [6], [8], [13], [14], [15], [16], [19] [20], [21], [22], [24], [26].

In examining the formalist drive-by it is useful to consider the claim that Whitehead and Russell ought to provide a precise presentation of the syntax of the system. That is, they ought to have provided a 'purely formal' description of the system of *Principia* 'in abstraction from' any interpretation of the system ([4]: 58); minimally, they should have characterised the 'syntax' of *Principia* in the narrower sense of specifying the formation and inference rules of the system in such terms. To illustrate what is at issue, Whitehead and Russell state the principle of inference generally known as modus ponens (for the propositional fragment of *Principia*) thus: '$*1 \cdot 1$. Anything implied by a true elementary proposition is true.' ([31]: 98). Jung amongst others takes Whitehead and Russell's primitive proposition ∗1·1 to reflect a failure to distinguish an object language under discussion (namely the ideal language of *Principia Mathematica*) from a metalanguage in which this object language is being described ([16]: 137-140). On this view, the inference rule in question should be formulated in the metalanguage in a form such as the following: 'for any formulae ϕ and ψ, from ϕ and $\ulcorner \psi \rightarrow \psi \urcorner$ infer ψ,' ([16]: 140, using corners for quasi-quotation and syntactical variables ranging over the well-formed formulas of the object language).

I note for future reference that Jung's objection to $*1 \cdot 1$ assumes that propositions are, in an appropriate sense, *Principia* formulas. It will be seen below that if *Principia* propositions are not formulas but other kinds of entities then the formalists have not demonstrated an error in the Whitehead-Russell presentation, but rather have failed to address the system under discussion. Before considering what *Principia*'s propositions might be however I consider two arguments in favour of the claim that the Whitehead-Russell presentation of *Principia* is defective in so far as it departs from the the the formalists approach. These arguments may be summarised thus:

Effectiveness requirement The Whitehead-Russell presentation, to the extent that it departs from the formalist approach, fails to provide a notion of proof that is effective, and hence the ideal language of *Principia* is not adequate for the purpose of communication (c.f. [4]: 50-52; [11]: 616 'Note added 28 August 1963');

Appropriate metatheory requirement A formalist reconstruction of the logic of *Principia* is to be preferred to the original Whitehead-Russell presentation since the metatheoretical properties of the former system may be precisely characterised for *at least* the first-order fragment, and moreover for this fragment the properties are well-known and appropriate - in brief, the theory is sound, complete, consistent and undecidable.

Consider firstly the effectiveness requirement. This claim is not essential to the formalists' position, as the idea of a 'formal system' may be defined more broadly so as to avoid this requirement (e.g. [23]: 25). That is as well for the formalist perspective, as the routine communication of working mathematicians presents strong prima facie evidence against this claim. To illustrate this point in various ways I will assume that the currently accepted metatheory of (classical) first-order logic and so on is essentially correct. Given this assumption, the doctrine in question implies the absurd claim that the language of informal arithmetic is not adequate for communication and yet that the (formal) language of a (standard) first-order theory of arithmetic (such as Mendelson's S [23]: Chapter three) is adequate for communication (even though the theory is, by well-known arguments stemming ultimately from [11], undecidable if ω-consistent). As another example, by the formalists account one cannot effectively determine whether an arbitrary formula of the restricted functional calculus is or is not a theorem (assuming the system exhibits a property similar to consistency [5]); and yet we are to believe that first-order languages are adequate for communication (including the notion of logical consequence). Yet if language users must (to communicate) be provided with effectively specified rules for identifying a formula and so on, how is it that they are able to communicate when theoremhood is only semi-decidable? These informal observations are not decisive; a formalist might in principle accept them and yet maintain that the effectiveness requirement is reasonable. To defend a criticism of *Principia* in these terms however some evidence in support of the claim is needed (including an explanation of the behaviour of, for example, working mathematicians who appear to be able to communicate effectively using the language of informal arithmetic).

I turn now to the appropriate metatheory requirement. The issue I wish to consider is whether the formalist metatheory of (classical) first-order logic can be made precise in the expected way. If not then it is question begging to rebuke Whitehead and Russell for not following this approach. To simplify discussion, I focus on Mendelson's [23] account of the pure predicate calculus PP though the demonstration that follows applies, with appropriate changes, to any standard (classical) first-order logic.

2.1 Expressing the metatheory of PP in NBG

When the metatheory of first-order logic is informally presented the question arises as to whether the metatheory can be made precise in some way. For example, Cohen, writing for the general mathematician who is not a specialist in logic, asserts that this can be done as follows:

> Having now given rules for forming valid statements we come to the problem of identifying these statements with the intuitively "true" statements. This discussion will be carried out in the spirit

> of traditional mathematics, that is to say, outside of any formal language. We shall use some elementary notions of set theory. After we have formalised set theory itself, then of course this discussion can be expressed in that formal system. ([7]: 11-12.)

I will leave the required sense of 'expressed' undefined and for brevity I assume as given Mendelson's presentation of NBG set theory ([23]: Chapter Four). In view of the following it appears unlikely that the metatheory of first-order logic may be made precise in this sense.

Proposition 2.1. *If $\mathfrak{D}$ is the domain of a standard interpretation $\mathfrak{M}$ of the language of NBG set theory that is a model of NBG which expresses the metatheory of the pure predicate calculus PP then $\mathfrak{D}$ is and is not a class.*

Proof Sketch. I note firstly that $\mathfrak{M}$ is a standard interpretation of the language of NBG set theory if and only if $(A_2^2)^M$ is the membership relation defined on the domain of $\mathfrak{M}$. I note secondly that:

Proposition 2.2. *It follows from the definition (or specification) of a domain of interpretation of an (arbitrary) first-order language that an object is a domain of interpretation of the language of NBG set theory if and only if it is a domain of interpretation of the language of PP*

Proof. The proposition follows immediately from the observation that the definition/specification of the notion applied for the one language is the same as that applied for the other. □

Thus suppose for example that a domain of interpretation of a first-order language is defined or specified to be a non-empty set. Then if X is the class of all domains of interpretation of the language of NBG set theory and Y is the class of all domains of interpretation the language of PP then $X = Y$. For the proof of Proposition 2.1 I hereon assume that an object is a domain of interpretation of a first-order language if and only if it is a non-empty class, proper or otherwise ([23]: §2.2). Nothing essential rides on this as the following proof holds, as will be seen, if we assume alternatively that a domain is a class or a set (without qualification). For the proof of Proposition 2.1 I thus hereon only need to establish that $\mathfrak{D}$ is not a class (as we have, by hypothesis, that it is).

Throughout the following let L be the language of NBG set theory, $\mathfrak{D}$ be the domain of interpretation $\mathfrak{M}$ of L of interest (i.e. under $\mathfrak{M}$, L expresses the metatheory of PP). The proof that $\mathfrak{D}$ is not a class requires only exclusion of two cases: that $\mathfrak{D}$ is a set (improper class or NBG class that is a member of another NBG class) and that $\mathfrak{D}$ is a proper class. The proof is as follows:

$\mathfrak{D}$ is a set Assume to derive a contradiction that $\mathfrak{D}$ is a nonempty set.

1. Then $\mathfrak{D}$ contains at least every nonempty set since:
    - (by hypothesis) an object is a domain of interpretation of a first-order language if and only if it is a non-empty class (proper or otherwise);
    - under $\mathfrak{M}$, some sentences of L quantify over all domains;
    - hence every non-empty set must be in $\mathfrak{D}$ (and moreover every non-empty proper class as well) since the only quantified variables occurring in L are individual variables, and, under the assumed standard (Tarskian) semantics, an object is in the range of a quantified individual variable under $\mathfrak{M}$ only if it is in the domain $\mathfrak{D}$ of $\mathfrak{M}$;
    - thus to avoid the contradiction that some non-empty class in $\mathfrak{D}$ is both a proper class and not a proper class (being a member of $\mathfrak{D}$) we must assume that every non-empty class is a set;
2. But Item 1 yields a contradiction, since:
    - we can prove in NBG (by contradiction) that the collection of all sets V is not a set: if the universal class V were a set then by the Class Existence Theorem ([23]: Corollary 4.4) there would exist a class Y that included every element of V that was not a member of itself; but since Y is a subclass of a set V, Y must be a set (by [23]: Corollary 4.6b); thus $(\exists Z)(V \in Z)$ implies a contradiction, that $Y \in Y$ holds iff $Y \notin Y$ holds.
    - The proof that V is not a class (with appropriate changes) holds if we consider the class that is equal to $V - \emptyset$. Thus the assumption that $\mathfrak{D}$ is a set yields a contradiction.

$\mathfrak{D}$ is a proper class Assume to derive a contradiction that $\mathfrak{D}$ is a proper class. But then $\mathfrak{D}$ must contain $\mathfrak{D}$ itself since (as just observed):

- under $\mathfrak{M}$, some sentences of L quantify over all domains;
- since $\mathfrak{D}$ itself is a domain of interpretation of L this implies, by Proposition 2.2, that under $\mathfrak{M}$, $\mathfrak{D}$ itself is in the range of some quantified NBG variable occurring in some L sentences;
- but this implies that $\mathfrak{D}$ itself is a member of the domain $\mathfrak{D}$ of $\mathfrak{M}$ - since the only quantified variables occurring in L are individual variables, and, under the assumed standard (Tarskian) semantics,

> an object is in the range of a quantified individual variable under $\mathfrak{M}$ only if it is in the domain $\mathfrak{D}$ of $\mathfrak{M}$.

But this implies the contradiction that $\mathfrak{D}$ is both a proper class (by hypothesis) and not a proper class (being a member of the class $\mathfrak{D}$).

□

Proposition 2.1 suggests that the formalist metatheory of first-order logic is in some sense false; that, for example, when the metatheory is made precise in some way a contradiction arises. To determine whether Proposition 2.1 supports this broader conclusion, one must also consider whether the problem that arises in attempting to make the metatheory precise in the expected way:

- is avoided under a non-standard interpretation that is a model of NBG set theory;
- is due to some peculiarity of NBG that does not affect some alternative set theory or non-set theory based approach to formalising the metatheory of first-order logic;
- can be avoided by adopting some non-standard semantics for NBG set theory;
- might be avoided by assuming that the logic of the metatheory is in some sense non-classical.

While the determination of the broader significance of Proposition 2.1 in this sense is beyond the scope of what is attempted here two observations are relevant. Firstly, the formalist metatheory of first-order logic must be false if our informal metatheoretic reasoning about this object of inquiry (the metatheory of first-order logic) breaks, legitimately, from either the law of contradiction or the law of excluded middle. Thus we may ignore the fourth point above in considering whether Proposition 2.1 implies that the formalist metatheory is false. Secondly, it can be shown that the formalist metatheory of first-order arithmetic is incompatible with classical arithmetic in the following sense: to avoid the conclusion that this metatheory is subject to paradox the formalist must assert that there exists a theorem of formalist, first-order arithmetic that is false under the standard interpretation [2]. Thus faced with a choice of rejecting either informal classical arithmetic or the formalist metatheory of first-order arithmetic it most classical logicians will reject the latter. Nevertheless, the contrary view may be dealt with as follows. Consider the claim that in light of this result, informal classical arithmetic should be rejected in favour of of the formalist metatheory of first-order arithmetic. The claim is self-defeating in the sense that given this hypothesis, we may turn the usual argument based on Gödel's arithmetisation of the syntax of such systems on its head: let $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_n$, ... be an be an enumeration of the theorems of the system that are, by hypothesis, false under the standard interpretation M; to show that the informal metatheory is not false, we must exclude the possibility that we may arrange our arithmetisation of the syntax of this system so that some (informal) metatheoretical statement of a property of the system goes over to one of these formulas. For: (i) the informal argument that the metatheory of the system is meaningful - and in particular that, the Gödel sentences are 'meaningful' propositions - is based on an appeal to the intuition that informal propositions concerning the (recursive) arithmetic of natural numbers are always 'meaningful' (e.g. [10]: 21); and (ii) the intended interpretation of the (first-order) formal system that expresses the arithmetised syntax is clearly the domain of classical arithmetic itself, so that the whole construction now gives reason to doubt the informal metatheory, rather than grounds for affirming its correctness as regards content.

For the remainder of this paper however I will (for the most part) put aside the question of the broader significance of Proposition 2.1. My aim instead is to expose a distinct though related weakness in the formalist critique of *Principia*, namely: the formalists assert that Whitehead and Russell are in error in so far as their description of the logic of *Principia* diverges from the contemporary (formalist) approach to describing logical systems; yet even if even if the formalist metatheory can be made precise in a way that avoids the problem identified in Proposition 2.1, this does not establish that the formalist approach to describing first-order logic is the only successful approach. To establish this point I turn now to considering the idea that *Principia* presents a successful alternative approach. The sketch of the first-order fragment of *Principia* suggests that *Principia* may avoid the above problem since the metatheory of this fragment does not appear to require the claim that there exists a well-defined totality of interpretations of the first-order fragment *Principia*; indeed any statement about 'all' such interpretations is, by *Principia*'s much disputed vicious circle principle, 'meaningless' in the required sense.

3 The first-order logic of *Principia*

In this section I assume as given Whitehead and Russell's description of the first-order logic of *Principia* and focus on the question of what propositions and propositional functions might be. Whitehead and Russell's account of first-order logic makes use of certain rules (e.g. substitution rules) that are not explicitly stated in *Principia.* I put aside such issues in the following discussion since any formulation of the the required rules depends upon a determination of what propositions and propositional functions (of different types / orders) are.

Although Whitehead and Russell provide informal explanations of their notions of a propositions / propositional functions in *Principia*, the notions are primitive or undefined in their system. In formalist reconstructions of *Principia* it is commonly assumed that propositions are either symbolic entities (such as well-formed formulas [16]) or the semantic entities associated with certain symbolic entities under some interpretation (such as facts [22] or attributes for propositional functions). Given either assumption for propositions (or, with appropriate changes, propositional functions) it appears that the discussion of propositions / propositional functions in *Principia* and related publications is essentially confused [26]. Thus if one assumes that propositions / propositional functions must be such entities then it is a forgone conclusion that Whitehead and Russell fail to describe a system of logic in *Principia.* If we wish to determine whether Whitehead and Russell describe a system of logic in *Principia* we must consider what propositions and propositional functions must be for Whitehead and Russell to have succeeded. One option which I sketch below is that (elementary) propositions are essentially ordered pairs $<\mathrm{p}, F_h(\mathrm{p})>$ of *Principia* symbol(s) p for elementary propositions associated with the 'meanings' $F_h(\mathrm{p})$ assigned to these symbol(s) under certain interpretations (set out below). The idea that a proposition is a sentence in association with its 'meaning' is not of course new; Church suggests that the idea is several hundred years old ([4]: 26). For reasons just stated, it is important to consider the possibility that propositions are such entities in evaluating the formalist criticism of *Principia.*

In describing the first-order logic of *Principia* I will firstly examine *Principia*'s propositional fragment or 'theory of deduction'. While the focus is on the system of the first edition of *Principia* I will make use of ideas presented in the *Introduction* to the second edition without thereby endorsing the revisions Russell proposes therein (or elsewhere in the second edition). To simplify presentation I will generally omit details concerning the handling of dots for parentheses.

3.1 The Theory of Deduction

My aim in this subsection is to indicate how *Principia*'s notions of an (elementary) proposition and an (elementary) propositional function may be made precise. I begin with a description of the semantic entities or 'meanings' that are to be associated with *Principia* symbol(s) for an elementary propositions / propositional functions. Although the following discussion is informal, it should be clear that, once the required interpretation is described, the metatheory for, say, the first-order fragment of *Principia* is expressed, under an appropriate interpretation, in a higher order fragment of *Principia* itself. When reasoning informally I will (to speak roughly) use $\vee$, $\overline{A}$, &, $\leftarrow$, $\leftrightarrow$, (Ey), (x), $\{x|F(x)\}$ for (respectively): disjunction, negation, conjunction, the (material or formal) conditional, the (material or formal) biconditional, existential and universal quantification, and ⌜the class of objects x such that $F(x)$ holds⌝. For brevity I avoid detailed discussion of type-theoretic complications.

Let an interpretation or (set-theoretic) structure for the propositional fragment of *Principia* be an ordered pair $< M, F >$ of a non-empty collection of objects of some type (the domain M) together with a function F from the well-formed formulas of the propositional fragment of *Principia* into M. (*Principia*'s class theory does not, strictly speaking, allow for such mixed classes. It may be seen below however that the required entities can be precisely described within *Principia*'s approach, since we can make typically ambiguous statements about propositional functions of a certain form, and propositional functions may relate terms of different types.) The truth assignment [1] or evaluation of the set of propositional formulas obtainable from a set of atomic or prime formulas is the usual focus of interest (which Whitehead and Russell hint at in passing ([32]: 115) and Post explores in detail [25]). For the truth assignment interpretation however the choice of elements of M it is somewhat arbitrary (one might use '+' or '-', or '0' or '1' etc) whereas for the interpretation considered below the nature of the elements of M is relevant. To introduce these elements I adapt an ontology sketched in [32] (c.f. [16]). According to this ontology, the universe consists of (i) individuals (which are neither propositions nor propositional functions) having various (ii) properties and standing in various relations, and of (iii) propositions and propositional functions of various types which may (truly or falsely) be predicated of such things (and of propositions and propositional functions of lower types) and (iv) of variables that (informally) range over entities of some given type. To illustrate, consider the elementary proposition that the individuals a and b stand in the relation R ([32]: 43). This proposition is true if and only if the complex of a and b standing in the relation R exists.

In describing a class of interpretations of interest the following preliminary definitions are used.

1. Let B_0 be the smallest set that contains every elementary complex that

exists. To clarify, I note that if the individuals a and b do not stand in the relation R then there exists the elementary complex $\overline{R}(a, b)$ of these individuals not standing in this relation.

2. For brevity, if $x \in B_0$ is the complex $T(x_1, \ldots, x_n)$ (or $\overline{T}(x_1, \ldots, x_n)$) that consists of the n-tuple of individuals $x_1, \ldots, x_n$ standing (or not standing respectively) in the relation T, let $\nu(x)$ be the elementary complex $\overline{T}(x_1, \ldots, x_n)$ (or $T(x_1, \ldots, x_n)$ respectively). Clearly if the law of contradiction holds then only one of $\overline{T}(x_1, \ldots, x_n)$ or $T(x_1, \ldots, x_n)$ exists and if the law of excluded middle holds at least one of these exists.

3. Let B_1 be the smallest set that contains every ordered pair $< x, y >$ such that: $x \in B_0$ & $(y = x$ or $y = \nu(x))$. In the case that $x \neq y$ case, the elementary complex $\nu(x)$ does not exist; while this may be dealt with precisely using Russells theory of descriptions the details are omitted here. Where convenient I assume that any element of B_1 of the form $< x, \nu(x) >$ is equal to $< x, \emptyset >$, since this simplifies discussion (though I avoid the definition required since it is less intuitive than the above).

4. Let ϕ be the function with domain B_1 such that:

$$\phi(< x, y >) = \begin{cases} < x, x > & \text{if } y = \nu(x), \\ < x, \nu(x) > & \text{if } x = y. \end{cases} \tag{1}$$

5. Let $\mathcal{B}_2$ be set that results from the power set of B_1 when the empty set is removed: $\mathcal{B}_2 = \wp(B_1) - \emptyset$.

6. Let $\mathcal{B}_3$ the power set of $\mathcal{B}_2$: $\mathcal{B}_3 = \wp(\mathcal{B}_2)$.

7. With $\{A_i\}_{i \in I}$, for $A_i \in \mathcal{B}_2$, let:

   - $\dot{\wp}\{A_i\}_{i \in I}$ be the smallest set that includes every set that contains at least one element of each member of $\{A_i\}_{i \in I}$.
   - $\Upsilon(\{A_i\}_{i \in I})$ be the set that results from $\{A_i\}_{i \in I}$ when, for all $i \in I$, each u in A_i is replaced by $\phi(u)$.

8. Let $\mathcal{H}$ be the class of all non-empty, total functions h with the class of *Principia* variables for an elementary proposition as domain and $\mathcal{B}_3$ as range.

When the above definitions are made precise $\mathcal{B}_3$ is a class in *Principia*'s sense. $\mathcal{B}_3$ will be the domain M of the interpretations of interest defined below. To describe the interpretations of interest the class of functions F_h defined as follows is required.

Definition 3.1. *Given a fixed but arbitrary choice of* h *in* $\mathcal{H}$*, we may define a function* F_h *by induction on (the length of)* p *a well-formed-formula of the propositional fragment of* Principia *as follows:*

- *Base case (*p *is a* Principia *variable for an elementary proposition):* $F_h(\mathrm{p}) = h(\mathrm{p})$

- *Induction step (for any formulas* $\mathrm{q}, \mathrm{r} < \mathrm{p}$*): given* $F_h(\mathrm{q}) = \{A_i\}_{i\in I}$ *and* $F_h(\mathrm{r}) = \{A_j\}_{j\in J}$*, with* $A_i, A_j \in \mathcal{B}_2$*:*

  If p **is** $\ulcorner \mathrm{q} \vee \mathrm{r} \urcorner$**:** $F_h(\mathrm{p})$ *is* $\{A_i\}_{i\in I} \cup \{A_j\}_{j\in J}$

  If p **is** $\ulcorner \sim \mathrm{q} \urcorner$**:** $F_h(\mathrm{p})$ *is*

$$\begin{cases} \Upsilon[\dot{\wp}\{A_i\}_{i\in I}]\, \mathit{if} \{A_i\} \neq \{A_1\} \\ \{y|(Ez)[z \ \in A_1 \ \& \ y = \{\phi(z)\}]\} \quad \mathit{otherwise} \end{cases}$$

For brevity, these functions F_h are sometimes referred to as 'interpretation functions' below. The definition of F_h tacitly assumes that the axiom of choice holds for $\mathcal{B}_2$. It will be evident shortly that, in light of the proposed definition of an elementary proposition, this assumption is required if axiom *1.7 is to hold: 'If p is an elementary proposition, $\sim p$ is an elementary proposition. Pp' [32].

For any choice of h in $\mathcal{H}$, the ordered pair of M and the associated function F_h thus constitutes an interpretation $\mathfrak{M}_h$ of the propositional fragment of *Principia.* Let $\mathcal{F}$ and $\mathcal{M}$ be the class of all these functions F_h and associated interpretations $< M, F_h >$ respectively. The required notion of a proposition may now be defined. Let p be some well-formed formula of the propositional fragment of *Principia*, and for $\mathfrak{M}_h$ in $\mathcal{M}$ let F_h be the associated member of $\mathcal{F}$. The proposition p shall be the ordered pair $< \mathrm{p}, F_h(\mathrm{p}) >$ for some definite choice of F_h in $\mathcal{F}$. In discussing the proposition p it is the ordered pair $< \mathrm{p}, F_h(\mathrm{p}) >$ that is mentioned and not simply the formula p nor the 'meaning' $F_h(\mathrm{p})$ assigned to this formula under $\mathfrak{M}_h$. Thus propositions in the ontology of *Principia* exhibit a dual nature, comprising both a symbolic component and a non-symbolic component which in the case of true elementary propositions consists of the entities the sentence that expresses the proposition are about. In the language of *Principia*, the symbols which express propositions are 'incomplete' and the objects of a single judgement are plural ([32]: 43). This feature of *Principia*'s propositions has caused a lot of confusion, particularly in relation to the theory of logical types discussed further below. At this point however I will simply recap some more basic features of *Principia*'s account of propositions.

I note firstly that if p is a *Principia* symbol for an elementary proposition, p is, in appropriate metalinguistic contexts, a variable whose value ranges over the value of every elementary proposition; that is, as the choice of F_h is varied over every member of $\mathcal{F}$, the value of $F_h(\mathrm{p})$ varies over the 'meaning' of every elementary proposition. It may also be seen that metalinguistic analogues for the usual notions of a (truth-functional) tautology and contradiction (for elementary propositions) may be defined using the class of all interpretations in $\mathcal{M}$. To see this observe firstly that the idea that an elementary proposition p (or $< \mathrm{p}, F_h(\mathrm{p}) >$) is true may be defined as follows:

Definition 3.2. p *is true under* $\mathfrak{M}_h$ *(or the proposition* $< \mathrm{p}, F_h(\mathrm{p}) >$ *is true) iff* $F_h(\mathrm{p})$ *contains a set* X *such that for every ordered pair* $< x, y >$ *in* X $x = y$.

The idea that an elementary proposition p (or $< \mathrm{p}, F_h(\mathrm{p}) >$) is a tautology or contradiction may then be defined as follows:

Definition 3.3. *An elementary proposition* p *(or* $< \mathrm{p}, F_h(\mathrm{p}) >$*) is a tautology (or contradiction) iff for all* f *in* $\mathcal{F}$*, the proposition* $< \mathrm{p}, f(\mathrm{p}) >$ *is true (or false respectively).*

Since the class of all interpretations in $\mathcal{M}$ is a well-defined totality from the perspective of *Principia*'s type theory, the notions of (propositional) logical truth and so on defined in these terms should not give rise to any contradictions. I note in passing that, at the cost of introducing cumbersome circumlocutions, we could simplify the ontology associated with the interpretation functions $F_h(\mathrm{p})$ and association notions as follows: associate with any such function the function F'_h which, roughly speaking, is equal to F_h for any well-formed formula p of the propositional fragment except that (i) every set X in $F_h(\mathrm{p})$ such that for every ordered pair $< x, y >$ in X $x = y$, is replaced by the corresponding set X' of entities x; and (ii) every set X in $F_h(\mathrm{p})$ such that for some ordered pair $< x, y >$ in X $x \neq y$, is deleted (so that where every set in $F_h(\mathrm{p})$ is deleted then the value of $F_h(\mathrm{p})$ is not defined). Thus the account could be developed, in the direction that Russell may have intended, such that a mapping of symbols to 'bits' of the real world are associated with true (elementary) propositions while false propositions are 'meaningful' only in the sense that the symbols are in the domain of the (revised) interpretation functions. I turn now to consideration of the first-order fragment of *Principia*.

3.2 The Theory of Apparent Variables

In this subsection I sketch an extension of the above approach to the first-order fragment of *Principia*. To simplify, I consider only propositional functions involving only non-constant symbols for either elementary propositions (p, q, $\ldots$) or (arbitrary values of) monadic propositional functions (ϕx, ψx, $\ldots$, ϕy, ψy, $\ldots$). Where convenient I will refer to this as simply 'the first-order fragment' of *Principia*, though this is not strictly the full first-order fragment. For brevity I assume as given Whitehead and Russell's description of this fragment of *Principia* and continue to put aside explicit consideration of the type theory of the system, formulation of omitted substitution rules and so on.

The question at hand then is this: what might a first-order proposition or propositional function be, if the Whitehead-Russell account describes a system of first-order logic? Let an interpretation or (set-theoretic) structure for the first-order fragment of *Principia* be an ordered pair $< M, F >$ of a non-empty collection of objects of some type (the domain M) together with a function F from the well-formed formulas of the first-order fragment of *Principia*, PM_1, into M. (For brevity I take the notion of a well-formed formula of the first-order fragment of *Principia* as given, though the idea can clearly be made precise in orthodox terms.) I begin by describing a class of interpretations of interest which all have as domain M the set $\mathcal{B}_3$ defined above. To describe the function $F_{\theta(\xi)}$ mapping members of PM_1 into M the following preliminary definitions are used. (For brevity I will sometimes, when no confusion should result, blur the distinction between the letter for some entity and the entity itself; e.g. I may mention the occurrences of 'an individual variable' x in a formula when strictly speaking only letters for variables occur in (an instance of) a formula.)

1. Let Ξ be the class of all functions ξ mapping individual variables of *Principia* (x, y, $\ldots$) into the class of all individuals (a, b, $\ldots$).

2. Let Ψ be the class of all *Principia* symbols $\phi(x)$, $\psi(x)$, $\ldots$ for an arbitrary value of an elementary propositional function.

3. Let p(x) be a member of PM_1 in which one or more occurrences of the (letter for an) individual variable x are free, and y be an individual variable that is free for x in p(x) in the usual sense:

   (a) p[y|x] shall be the formula that results when the variable y is substituted for every free occurrence of x in p(x);

(b) p[_|x] shall be be the class defined as follows:

$$\text{p}[\_|\text{x}] = \{\text{q}|(E\text{z})(\text{z is free for x in p(x) \& q} = \text{p}[\text{z}|\text{x}])\} \qquad (2)$$

4. Let $\Theta(\xi)$ be the class of all functions θ_ξ mapping Ψ into $\mathcal{B}_3$ such that (with p(x) as above):

   (a) ξ(x) occurs in θ_ξ(p(x));

   (b) there exists a class $k[\theta_\xi(\text{p(x)})]$ of occurrences of ξ(x) in θ_ξ(p(x)) such that (for arbitrary q = p[z|x] in p[_|x]): $\theta_\xi(\text{q}) = \theta_\xi(\text{p(x)})$ except that ξ(z) occurs in θ_ξ(q) in place of ξ(x) iff this occurrence is in $k[\theta_\xi(\text{p(x)})]$.

5. Suppose that (for $A_i, A'_i \in \mathcal{B}_2$) $X = \{A_i\}_{i\in I}$ and $Y = \{A'_i\}_{i\in I}$, have (informally) the same 'structure' except that some corresponding ordered pairs of elementary complexes involved differ only with respect to the individuals involved. Then $X\dot{\cup}Y$ shall be the $\{A''_i\}_{i\in I}$ that results from forming the union of corresponding elements of X and Y. Given a collection $\{A^1_i\}_{i\in I}$, $\{A^2_i\}_{i\in I}$, ... of such sets, with elements indexed by J, $\dot{\cup}_{j\in J}\{A^j_i\}_{i\in I}$ is the obvious generalisation of this operation.

To describe the required sense mentioned at Item 4, note that any $m \in M$ has the form $[(\alpha, \ldots)\ldots]$ or $[(\alpha, \beta, \ldots), (\gamma, \delta, \ldots), \ldots]$ etc where α, β etc are ordered pairs. As both elements of each such ordered pair are elementary complexes, such as $R(a, b)$ or $\overline{R}(a, b)$, at least one of which exists, the individual a occurs in m in the required sense if a is a constituent of one of the elementary complexes in $\bigcup m$ the union of all the sets in m. A class of functions required to define the notion of an an elementary (monadic) propositional function may then be defined as follows.

Definition 3.4. *For $\theta_\xi \in \Theta(\xi)$, define $F_{\theta(\xi)}(\text{p})$ by induction on* p *(a well-formed-formula of the first-order fragment of* Principia *that is* not *in the propositional fragment) as follows:*

- *Base case* (p $\in \Psi$)*: $F_{\theta(\xi)}(\text{p}) = \dot{\cup}_{\text{z}\in\text{p}[\_|\text{x}]}\theta_\xi(\text{z})$*

- *Induction step (for arbitrary* q, r < p*): given $F_{\theta(\xi)}(\text{q}) = \{A_i\}_{i\in I}$ and $F_{\theta(\xi)}(\text{r}) = \{A_j\}_{j\in J}$, with $A_i, A_j \in \mathcal{B}_2$:*

  If p **is** $\ulcorner$q $\vee$ r$\urcorner$ **or** $\ulcorner\sim$ q$\urcorner$**:** *definition follows $F_h(\text{p})$*

  If p **is** $\ulcorner$(x)q$\urcorner$**:** *$F_{\theta(\xi)}(\text{p})$ is $\dot{\cup}_{\text{z}\in\text{q}[\_|\text{x}]}\theta_\xi(\text{z})$*

If p **is** $\ulcorner(\exists \mathrm{x})\mathrm{q}\urcorner$**:**

$$F_{\theta(\xi)}(\mathrm{p}) = \{A_k|(E\mathrm{r})[\mathrm{r} \in \mathrm{q}[\_|\mathrm{x}] \ \& \ A_k \in F_{\theta(\xi)}(\mathrm{r})]\}$$

The extension of the above method to obtain the required definitions of an elementary (monadic) propositional function may then be obtained by defining a further class of functions $F(\mathrm{p})$, by combining the values of $F_{\theta(\xi)}(\mathrm{q})$ and $F_h(\mathrm{r})$ for appropriate parts q and r of an arbitrary p in PM_1. The class of true (monadic), first-order propositions p (or $< \mathrm{p}, F(\mathrm{p}) >$) and first-order logical truths may then be defined as above (Definitions 3.2, 3.3), with appropriate changes. Since the class of all interpretations involved is a well-defined totality from the perspective of *Principia*'s type theory, the resulting notions of first-order logical truth and so on should not give rise to any contradictions.

It is important to note that the semantic concepts defined in this section are intended primarily to provide an informal explanation of various notions assumed as primitive in the system of *Principia*. In expressing these notions within *Principia* itself the assumption of the existence of certain entities (to be associated with *Principia* expressions) will sometimes require the hypothesis that the axiom of choice holds for certain sets. This apparently must be assumed (in the metatheory) as an additional hypothesis when required.

The approach sketched above is of course not the only, or even necessarily the best, such account. For example, to simplify discussion the above account blurs the distinction between the meaning of a symbol for the arbitrary value of a propositional function (e.g. ϕx) and the meaning of the sentence used in affirming all values of this function (e.g. $(x).\phi x$); it may well be the case however that a semantics for *Principia* which respects the distinction between 'any' and 'all' is to be preferred. While the focus of this paper is on *Principia*'s first-order fragment the following section briefly considers some additional issues that arises in defending the full-system of *Principia*. While a defence of the full system is beyond the scope of what is attempted here, it is possible to show that certain commonly repeated arguments against the full-system stand in need of reconsideration in the light of the above; thus while a focus on the first-order fragment falls short of a defence of the full system it represents a step in that direction.

4 A Glimpse of the full system of *Principia*

In sketching an approach to extending the above account to the full system of *Principia* my aim is to shed some light on the following key questions about the full system: What is the type of a proposition / propositional function and how does that relate to the order of a proposition / propositional function? Is there any merit to Russell's distinction between 'for any' and 'for all' (defended in the first edition but subsequently abandoned)? Is it possible to state the 'vicious circle principle' without contradicting this doctrine itself? Does Gödel's technique of the arithmetisation of syntax apply to the system of *Principia* and is the arithmetic of the system therefore subject to the error that affects the first-order formalist account [2]? My discussion of each of these questions below is restricted to briefly indicating how the account presented above may be integrated into a more detailed exposition of the full system of *Principia*.

Principia's theory of logical types is notoriously complicated, even when attention is restricted to the account presented in the foundational exposition of mathematical logic ([31]: Part I). One indication of this is the fact that the extensive secondary literature on this topic presents a variety of apparently inconsistent accounts on the theory; [22] provides a good survey of material to around that date, though *Principia* is, for various reasons, the focus of some more recent research. In the discussion above (§3) a number of complications discussed in this literature were avoided by imposing two restrictions: considering only propositional functions involving only non-constant symbols for either elementary propositions (p, q, $\ldots$) or (arbitrary values of) monadic propositional functions (ϕx, ψx, $\ldots$, ϕy, ψy, $\ldots$) and focusing on the first-order case. The first difficulty to be addressed in relating this account to the broader literature is that it appears to conflict with Whitehead and Russell's theory in certain respects. The following two quotations illustrate the point of apparent conflict:

> For reasons explained in Chapter II of the Introduction, it would seem that negation and disjunction and their derivatives must have a different meaning when applied to elementary propositions from that which they have when applied to such propositions as $(x).\phi x$ or $(\exists x).\phi x$. [31]: 133

> We will give the name of *first-order propositions* to such as contain one or more apparent variables whose possible values are individuals, but contain no other apparent variables. First-order propositions are not all of the same type, since, as was explained in $*9$, two propositions which do not contain the same number of apparent variables cannot be of the same type. [31]: 169

The above account does in fact differ from Whitehead and Russell's own version in some important respects highlighted in these quotes. In the course of the following discussion I aim to show that: the differences do not involve any breach with Russell's vicious circle principle, but rather arise from implementing this principle differently for the first-order case. To explain this point however I will need to meander through a discussion of various notions involved, in the course of which some of the other questions set to be answered in this section will also be addressed.

It is firstly necessary to clarify the notion of a propositional function $\phi\hat{x}$ of a definite type (say n). From §3 it should be clear that $\phi\hat{x}$ is, on the account proposed, a function that maps or associates arguments suitable to x to propositions of a definite type - one complication involved however being that in most contexts 'the propositional function' is not the function per se but rather an ordered pair ($< \phi\hat{x}, F(\phi\hat{x}) >$) of a *Principia* expression or symbol for this function, say '$\phi\hat{x}$', associated with the required function that is given as the value of a suitable interpretation function F that maps particles of (some fragment of) *Principia* into an appropriately defined collection of some kind. Indeed the complication is involved twice over, since the arguments to the propositional function $< \phi\hat{x}, F(\phi\hat{x}) >$ are also ordered pairs $< x, G(x) >$ of *Principia* symbols for some variable 'x' associated via a suitable function G with an appropriate kind of object $G(x)$. Thus in talking of the type of a variable it is the type of the ordered pair that is specified or defined, as in the case of variables for elementary propositions discussed above.

Having clarified the idea of a propositional function of a definite type, it is important to clarify the proposed notion of a 'typically ambiguous' propositional function; as will be clear, the treatment of a typically ambiguous proposition, variable and so on is essentially similar. For the notion of a 'typically ambiguous' propositional function proposed here it is essential to stipulate that *Principia*'s primitive symbols include both typically definite symbols (distinguished by, for example, suppressed type indices) and typically ambiguous symbols for each kind of symbol required - such as propositions (p, q, ...), logical constants ($\vee$, $\sim$, dots, etc), for propositional functions ($\phi\hat{x}$ etc) and for arbitrary values of these (ϕx etc). For reasons to be explained shortly, it is essential to recognise that these symbols are object language symbols, not syntactical or metalinguistic variables or formula schemata. On this approach, proofs of propositions involving typically ambiguous propositional functions are proofs of genuine *Principia* theorems. To see how it is possible that *Principia* may have such symbols, and why it is necessary, Russell's distinction between 'for any' and 'for all' must be discussed.

To begin with an illustration, let $\phi\hat{x}$ be a typically definite propositional function. The assertion of any value of this function is the assertion of some definite but unspecified one of its values ϕx. The assertion of all values of the function, $(x)\phi x$, by contrast is the assertion that ϕx is always true, where the only (internal) limitation on the range of values of the variable x is that given by the type of this variable. In [2] I illustrate how to make the idea of 'any numeral' precise in the setting of Tarskian semantics for first-order number theory: we may add a primitive symbol and assign semantic rules such that the meaning of this symbol is like a parameter with values restricted to the objects assigned to the numerals. The usefulness of the idea may be somewhat obscured in this case by the incidental fact that the system described there is inconsistent, so it may prove useful to adapt the illustration for the case of *Principia.* Consider the Definition 3.4 of $F_{\theta(\xi)}(\mathrm{p})$ for the base case; to be specific, consider any value ϕx of the propositional function $\phi\hat{x}$. For an alternative to the approach set out above we might specify that: (i) $F_{\theta(\xi)}(\phi x) = \theta_\xi(\phi z)$, where z is a definite but unspecified choice of an individual variable that is free for x in ϕx; and (ii) $< \phi x, F_{\theta(\xi)}(\phi x) >$ is true under $F_{\theta(\xi)}$ iff, $F_{\theta(\xi)}(\phi x)$ contains a set X such that for every ordered pair $< x, y >$ in X $x = y$. This illustrates the idea of a definite but unspecified value of the propositional function $\phi\hat{x}$, though it is not an analysis of the idea since the idea of 'any' is itself is used in the illustration.

Considered in its most generally setting, namely the context of a discussion of any value of a typically ambiguous propositional function, the idea of 'any value' is (by the vicious circle principle) an unanalysable primitive notion: no meaningful proposition or finite conjunction of propositions can specify the notion involved, since any definite proposition must have some fixed type (and so possible higher types of the ambiguous propositional function can not be mentioned in the proposition in question). For example, Whitehead and Russell's definition of the primitive idea of "being of the same type" ($*9\cdot131$) must be a typically ambiguous propositional function in this sense or else, by the vicious circle principle, the (open) sentence is meaningless: the definition involves the phrase "We say that u and v are of the same type ..."; clearly, if the definition is to cover any of the required cases we cannot fix the type of either the variables u and v or the notion of type involved - but if the definition is to be meaningful, then the types involved must be some definite but unspecified case. In symbols, the assertion of $*9\cdot131$ is thus the assertion of any value of a typically ambiguous propositional function $\psi(\hat{u}, \hat{v})$; in considering any value of this function of some definite type, we may infer the truth of $\psi(u, v)$ from $\vdash \psi(\hat{u}, \hat{v})$.

The terminology and notation involved may seem a little cumbersome but the price is well worth paying as we obtain a solution to the problem that otherwise the vicious circle principle (applied to propositions) appears to rule out the possibility of any truly general system of logic (as Russell noted in his initial presentation of the theory):

> The first difficulty that confronts us is as to the fundamental principles of logic known under the quaint name of “laws of thought”, “All propositions are either true or false”, for example, has become meaningless. If it were significant it would be a proposition and would come under its own scope. Nevertheless, some substitute must be found, or all general accounts of deduction become impossible. [27]

Alternatively, one might argue that there is something wrong with the vicious circle principle itself (from a classical point of view) in the sense that, as a number of critics have claimed, any attempt to state this principle must either contradict this principle itself and therefore either be itself a false statement (if the principle is true) or show that the principle itself is false (since it has such exceptions):

> ... it seems impossible to formulate the theory without violating its own provisions because words like ’function’, ’entity’, and ’type’ must always remain free from type restrictions. If, for example, we say that no function may be asserted significantly of all entities without distinction of type, our own statement involves the unlimited generality which it declares to be impossible. [17]: 670

The solution to both this objection, and to the associated problem identified by Russell in the above quotation, involves recognising that various logical terms (such as ’type’ and ’proposition’), are typically ambiguous and the required assertions involve assertion of any value of a typically ambiguous propositional function, not assertion of all values of some propositional function (which, being a definite object must have some fixed type and must therefore be beyond the scope of any variable it contains). (Incidentally, for emphasis note that the type of terms, on the proposed version of Russell’s theory, is a property of, roughly speaking, a symbolic entity associated with its ’meaning’, not a property of a purely symbolic entity such as a word, itself.)

On the account of type theory under discussion, Russell’s proposal in the second edition of *Principia* to abandon the distinction between ’for any’ and ’for all’ is, a serious error ([32]xiii). The elimination of this distinction requires much more than the rewriting of ’propositions as printed’ ([32]xiii) referred to by Russell. It requires, on the basis of the argument mentioned above, an abandonment of the theory of logical types. Another way of considering this problem, is to ask what ’happens’ to the metatheory of *Principia* if the distinction is abandoned? If we restrict attention to some specific fragment of the system (e.g. the first-order fragment for some definite type) then the statement of the axioms and rules of inference presents no special problem from the perspective of the theory of logical types. If we try to describe the full system of *Principia* however then we encounter the problem just indicated; that is, to define the notion of a type or, to state the rule of inference modus ponens (without restriction as to type) then our metatheory of *Principia* must either 'sit outside' of the entire system of *Principia* itself (so that the logic is not truly general) or the vicious circle principle must be false. The formalist strategy of trying to avoid this problem, by describing the inference rule as though only uninterpreted sequences of signs are considered, is not available since that strategy fails.

The theory of types proposed above, in other words, has the consequence that the full system of *Principia* contains its own metatheory and does so without any violation of the vicious circle principle (the assertion of which may be identified with the assertion of a suitable typically ambiguous propositional function). To avoid any confusion I will illustrate the application of the proposed account for the case of modus ponens (the case for the statement of the vicious circle principle is essentially the same). *Principia*'s statement of modus ponens may be identified with the following assertion of a typically ambiguous propositional function:

$*9{\cdot}12$ What is implied by a true premiss is true. Pp. [31]: 137

Thus in symbols, *Principia*'s rule of inference modus ponens is simply this :

$$\vdash \psi(\hat{u}, \hat{v}, \hat{w}) \tag{3}$$

The three terms involved are simply the two propositions or propositional functions, of some definite type, that are the premiss on the one hand and the conclusion on the other. It is important to note that the symbol sequence '$\psi(\hat{u}, \hat{v}, \hat{w})$' is a well formed-formula of *Principia*, not a metalinguistic syntactical expression or schema (since on the account presented here, *Principia*'s primitive symbols include typically ambiguous symbols as well as type specific symbols). In asserting any value $\psi(u, v, w)$ of $\psi(\hat{u}, \hat{v}, \hat{w})$ some definite (but unspecified) value of this typically ambiguous propositional function is asserted or said to be true; thus, a proposition of a definite type (say n) and an appropriate notion of 'truth' of a definite type are involved. Thus it is a proposition $\psi(u, v, w)$, not an uninterpreted formula, of a definite type that is asserted. The assertion $(u, v, w)\psi(u, v, w)$ of the rule for all propositions or propositional functions of a definite type n is of course a metatheoretical proposition in the sense that the type of the proposition $(u, v, w)\psi(u, v, w)$ must be greater than that of any proposition thus quantified over; and yet by exploiting the device of typical ambiguity, we may infer any value $\psi(u, v, w)$ of $\psi(\hat{u}, \hat{v}, \hat{w})$ from $\vdash \psi(\hat{u}, \hat{v}, \hat{w})$ without incurring any reflexive fallacies.

At this point I have presented the prerequisite material so that an explanation may be provided for the differences noted above between the type theory of §3 and the corresponding portions of *Principia*. Simply put, Whitehead and Russell propose that to assert $(x)\phi x$, with $\phi\hat{x}$ an elementary propositional function, one asserts, as noted above, that ϕx is always true (or true for any value of x of the appropriate type). Taking the idea of a propositional function as a primitive notion, Whitehead and Russell apparently infer from this that quantification over a collection of elementary propositions (ϕa, ϕb, etc, where a and b are individuals) is involved; thus the type of the first-order proposition $(x)\phi x$ must be greater than that of any of these elementary propositions. Generalising from this, they infer that, even though, for example, $(x)\phi x$ and $(x)(y)\phi(x, y)$ both be first-order propositions they must for this reason be of a different type. In the type theory for first-order logic of §3, by contrast, a (partial) analysis of the ideas of an elementary proposition and a first-order proposition (of the lowest such type) is presented - derived essentially from informal ideas sketched in *Principia.* On the basis of this analysis it appears that, for the first-order case of the lowest type, the process of generalising elementary propositions that produces the first-order propositions does not raise the type (since quantification over elementary propositions is not essentially involved); in other words, the assertion of §3 that the elementary proposition ϕa and the first-order proposition $(x)\phi x$ are of the same type, does not involve any breach with the vicious circle principle but rather results from a different analysis of the results of applying this principle.

While the above sketch of the full system of *Principia* is thin in many respects, there is nevertheless enough material to address the remaining question raised above, namely whether Gödel's technique of the arithmetisation of syntax actually applies to the system of *Principia.* Contemporary commentators, subject to the spell of Gödel's masterful proof [11], are apparently willing to concede this point on Gödel's recommendation without any discussion of details:

> As Gödel himself notes, however, his incompleteness proof only needs to invoke some fairly elementary features of the full-blooded theories of *Principia* and *ZF* ... So we can now largely forget about *Principia* ... [28]: 123

The cardinal arithmetic of *Principia* is developed in a (typically ambiguous) higher order fragment (following the plan of [27]: IX); thus an examination of *Principia*'s arithmetic is essentially beyond the scope of a study focusing on the first-order fragment. Nevertheless, from the above account of the formation and inference rules of *Principia* it is clear that Gödel's claim to have shown how to arithmetise the syntax of this system involves an implicit appeal to the formalist drive-by ([12]: 120) noted above (§2). That is, if we examine *Principia*'s rule of inference $*9 \cdot 12$ quoted above, it is clear that Gödel [11] does not intend to claim to have shown how to arithmetise an inference rule involving a primitive notion of 'truth'; rather, the claim is that if Whitehead and Russell where to correctly describe their system then the result would essentially correspond to system P, putting aside the introduction of proper axioms corresponding to Peano's postulates and other non-essential changes. As the formalist theory of arithmetic is however defective [2], the assertion that *Principia* ought to be adjusted in the way Gödel claims should be rejected; it has been shown above for the first-order case that the required notion truth can be made precise and it is not difficult to see that the required higher order notions can also be described. The system of *Principia* is thus resistant to Gödel's technique and therefore also avoids the defect that destroys the formalist theory of arithmetic [2].

The above sketch of the full system of *Principia* is lacking in details in a number of respects. For a fuller account, various issues not explicitly dealt with in the original Whitehead-Russell presentation, such as the substitution rules of the system, should be dealt with, though some of these issues have already been discussed to some extent in the existing literature (e.g. [16]).

5 Why *Principia* is a superior system

In light of the above, the widely accepted claim that the Whitehead-Russell presentation of logic in *Principia* is essentially defective in departing from contemporary standards should be rejected. Indeed, it is a *virtue* of the Whitehead-Russell presentation that it departs from these standards; the Whitehead-Russell version of classical first-order logic is closer to a gold-standard account than the currently accepted rivals discussed above. The evidence for this presented above may be summarised as follows:

- Firstly we have evidence (§2.1) suggesting intrinsic defects in the existing systems (which result from adherence to existing formalist standards). The orthodox metatheory of first-order logic involves propositions that generalise over all domains of interpretation of arbitrary first-order languages and so on, and one might reasonably doubt that paradox is avoided if this reasoning is conducted informally. Yet the method that is generally supposed to provide a precise statement of the metatheory (expression in a standard model of a first-order set theory such as NBG) fails. In view of the defects that affect the formalist theory of arithmetic [2] there is essentially no credibility to the claim that the formalist metatheory of first-order logic is viable.

- Secondly, we have evidence that that the first-order logic of *Principia* avoids the difficulties just mentioned (§3-§): The Whitehead-Russell presentation of first-order logic in *Principia* does not require the assumption that one can generalise about all interpretations of this fragment of *Principia.* (The contrary idea involves an obvious transgression of *Principia*'s type theory, though I have not considered such matters above.) The system of *Principia* resists Gödel's technique of arithmetisation and thus provides a viable classical theory of arithmetic.

If one accepts that *Principia*'s approach is to be preferred to the contemporary (formalist) rival account, it appears that revisions in a wide variety of notions basic to logic are implied. For example, *Principia*'s first-order notions of proposition, truth, logical truth and so, as sketched above, on do not correspond to the orthodox Gödel-Tarski concepts. *Principia*'s propositions for example are neither uninterpreted sequences of symbols nor the (purely) set-theoretic entities associated with these under some specific, broadly Tarskian, interpretation of the system; the elementary proposition p, to continue the example, is the ordered pair $< \mathrm{p}, F_h(\mathrm{p}) >$ (for some definite choice of F_h in $\mathcal{F}$). Whitehead and Russell's remarks within *Principia* itself about the nature of propositions (e.g. [32]: 44) are broadly consistent with this view. It is important to note that if the Whitehead-Russell account is correct then there is no purely syntactical characterisation of the formation and inference rules of first-order logic; thus it is begging the question to claim that *Principia* is defective in so far as the syntax of the system is not well-defined in this sense.

I turn now to a consideration of the light the above sheds on some debates concerning *Principia.* As noted above a number of commonly repeated objections to *Principia* beg the question as to whether the rival formalist account of logic should be preferred to the system of *Principia.* If one assumes, for example, that elementary propositions are either (exclusively) symbolic or non-symbolic entities then elementary propositions are not the kind of entities that *Principia* assumes or asserts them to be. Where the nature of propositions is at issue however, the bold assertion that that they must be either the one or the other is not a good criticism of *Principia*; some demonstration of the claim should also be provided. Quine's assertion that Russell fails to 'distinguish between propositional functions as notations and propositional functions as attributes and relations' ([26]: 152) is an example of this type of question- begging objection. A proof that propositional functions must be either the one or other kind of entity is nowhere hinted at in [26] but assumed as a given. Similarly, if relevant sections of *Principia* are read in light of the above then the often repeated claim that *Principia* exhibits a failure to distinguish use from mention appears to be false. While Whitehead and Russell's notion and terminology might be improved in this respect in some contexts, there is no sense in which a confusion of use and mention is inherent to the account of logic they present; quotations marks are often used appropriately in *Principia* (even if the conventions involved are not always transparent).

A similar kind of error appears to be involved in a more profound criticism of *Principia* deriving ultimately from Gödel's incompleteness theorems [11]. In an early exposition of this argument, Gödel [10] claims that *Principia*'s account of paradox must be wrong, since if we arithmetise the syntax of a suitable system one can:

> construct propositions which make statements about themselves, and, in fact, these are arithmetic propositions which involve only recursively defined functions and therefore are undoubtedly meaningful statements. [10]: 21.

In light of the above Gödel's criticism of *Principia*'s account of paradox fails on three points:

1. Gödel has not described a method of constructing 'propositions which make statements about themselves': if we assume *Principia*'s notion of a proposition as defined above then it appears that Gödel [11], [10] is rather discussing a proposition $< \mathrm{p}, F_{\theta(\xi)}(\mathrm{p}) >$ that makes a statement about the sentence or formula p (viewed as an uninterpreted sequence of signs) that expresses this proposition under a specific interpretation $[F_{\theta(\xi)}(\mathrm{p})]$. But this is not an example of a proposition $< \mathrm{p}, F_{\theta(\xi)}(\mathrm{p}) >$ that makes a statement about itself $< \mathrm{p}, F_{\theta(\xi)}(\mathrm{p}) >$. Gödel essentially notes this point himself when, in explaining the proof informally, he comments that:

   > Contrary to appearances, such a proposition involves no faulty circularity, for initially it [only] asserts that a certain well-defined formula . . . is unprovable. Only subsequently . . . does it turn out that this formula is precisely the one by which the proposition itself was expressed. [11]: 598, fn 15

2. In view of Proposition 2.1 and the failure of the formalist theory of arithmetic [2] the claim that paradox is avoided in systems subject to the Gödel phenomenon is no longer credible.

3. Gödel's claim that the phenomenon he describes 'is not in any way due to the special nature of the systems' ([11]: 597) is plainly false. The possibility of arithmetising the syntax of a system in the way that Gödel proposes requires that the formation and inference rules of the system conform to the formalist program. Ironically, it is the non-conformity of the system of *Principia* to these requirements, so openly deplored by Gödel ([12]: 120), that renders the system immune to the disaster that affects formalist arithmetic.

In summary, while *Principia* might be improved in some matters of detail it presents a system for classical logic that is superior in important respects to the currently accepted rival systems discussed herein.

References

[1] Jon Barwise, *An Introduction to First-order Logic*, Handbook of Mathematical Logic (Jon Barwise, ed.), North Holland Publishing, Amsterdam, 1977, pp. 5–46.

[2] Stephen Boyce, *On the Formalist Theory of Arithmetic*, arXiv:1010.1282v3 (2010).

[3] Charles S. Chihara, *Ontology and the Vicious-Circle Principle*, Cornell University Press, Ithaca, 1973.

[4] Alonzo Church, *Introduction to Mathematical Logic*, Princeton University Press, Princeton, 1956.

[5] Alonzo Church, *'A Note on the Entscheidungsproblem',* The Journal of Symbolic Logic I*, 40-41, 1936, corrections ibid 101-102*, in Davis [9], pp. 110–115.

[6] Alonzo Church, *Comparson of Russell's Resolution of the Semantical Antinomies with that of Tarski*, The Journal of Symbolic Logic **41** (1976), no. 4, 747–760.

[7] Paul J. Cohen, *Set Theory and the Continuum Hypothesis*, W.A. Benjamin, Amsterdam, 1966.

[8] Irving M. Copi, *The Theory of Logical Types*, Routledge & Kegan Paul, London, 1971.

[9] Martin Davis (ed.), *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions*, Raven, New York, 1965.

[10] Kurt Gödel, *'On Undecidable Propositions of Formal Mathematical Systems', mimeographed lecture notes, taken by S Kleene, J Rosser, 1934*, in Davis [9], pp. 39–74.

[11] Kurt Gödel, *'Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I',* Monatshefte für Mathematik und Physik ***38****, 173-198. English translation by J. van Heijenoort as 'On formally undecidable propositions of* Principia Mathematica *and related systems I', 1931*, in van Heijenoort [30], pp. 596–616.

[12] Kurt Gödel, *'Russell's Mathematical Logic', Reprinted from* The Philosophy of Bertrand Russell*, by Paul A. Schilpp. ed. Open Court Publishing, La Salle, Illionois, pp. 123-153*, Kurt Gödel: Collected Works, Volume III, Publications 1938-1974 (Solomon Feferman, John W. Dawson, Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, eds.), Oxford University Press, Oxford, 1999, pp. 119–141.

[13] Warren D. Goldfarb, *Logic in the Twenties: the Nature of the Quantifier*, The Journal of Symbolic Logic **44** (1979), no. 3, 351–368.

[14] Allen Hazen, *Predicative Logics,* The Journal of Symbolic Logic I*, 40-41, corrections ibid 101-102*, Handbook of Philosophical Logic, Volume 1: Elements of Classical Logic (D. Gabbay and F. Guenthner, eds.), D. Reidel, Dordrecht, 1983, pp. 331–407.

[15] A.P. Hazen and J.M. Davoren, *Russell's 1925 Logic*, Australasian Journal of Philosophy **78** (2000), no. 4, 534–556.

[16] Darryl Jung, *The Logic of Principia Mathematica*, Ph.D. thesis, Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 1994.

[17] William Kneale and Martha Kneale, *The Development of Logic*, Clarendon Press, Oxford, 1984.

[18] Gregory Landini, *Russell's Hidden Substitutional Theory*, Oxford University Press, Oxford, 1998.

[19] Gregory Landini, *Russell's Substitutional Theory*, The Cambridge Companion to Bertrand Russell (Nicholas Griffin, ed.), Cambridge University Press, Cambridge, 2003, pp. 241–285.

[20] Gregory Landini, *Quantification Theory in *8 of* ***Principia Mathematica*** *and the Empty Domain*, History and Philsoophy of Logic **26** (2005), no. 1, 47–59.

[21] Gregory Landini, *Russell's Schema, Not Priest's Inclosure*, History and Philsoophy of Logic **30** (2009), no. 2, 105–139.

[22] Bernard Linsky, *Russell's Metaphysical Logic*, CSLI Publications, Stanford, 1999.

[23] Elliott Mendelson, *Introduction to Mathematical Logic*, fifth ed., Chapman & Hall, London, 2010.

[24] John Myhill, *The Undefinability of the Set of Natural Numbers in the Ramified Principia*, Bertrand Russell's Philosophy (George Nakhnikian, ed.), Harper and Row, New York, 1974, pp. 19–27.

[25] E. Post, *'Introduction to a General Theory of Propositions'* American Journal of Mathematics ***43****, 163-185, 1921*, in van Heijenoort [30], pp. 264–283.

[26] W.V. Quine, *Introduction to B. Russell 'Mathematical logic as based on the theory of types'*, in van Heijenoort [30], pp. 150–152.

[27] B Russell, *'Mathematical logic as based on the theory of types',* American Journal of Mathematics 30*, 222-282*, in van Heijenoort [30], pp. 150–182.

[28] Peter Smith, *An Introduction to Gödel's Theorems*, Cambridge, Cambridge, 2007.

[29] Alasdair Urquhart, *The Theory of Types*, The Cambridge Companion to Bertrand Russell (Nicholas Griffin, ed.), Cambridge University Press, Cambridge, 2003, pp. 286–309.

[30] Jean van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic*, Harvard University Press, Cambridge, Mass., 1967.

[31] A. Whitehead and B. Russell, *Principia Mathematica*, first ed., vol. One, Cambridge University Press, Cambridge, 1910.

[32] A. Whitehead and B. Russell, *Principia Mathematica to *56*, abridged version of second ed., vol. One, Cambridge University Press, Cambridge, 1962.